

Closed-Loop Data Gap Mitigation for Safe AI Perception

Riken Mangukiya, Christian Freystein, Xinyi Li,
Anna Bieberstein, Dr. Xinxing Wang,
Jochen Schwenninger (Bertrandt Group)



Data Exporter

The Data Exporter analyzes annotated driving datasets according to scenario descriptions and data requirements. It gives a data subset by filtering ASAM OpenLABEL data using JSON configuration files and reports the matched frames and matched objects. The postprocessing scripts helps to identify whether the available dataset sufficiently covers the required Operational Design Domain (ODD) or whether data gaps exists.

In addition, the tool enables the selection of data subsets with specific characteristics, which can be used for targeted model training. The output of the Data Exporter provides information on the number of remaining samples, which serves as an

indicator of how well the defined data requirements are covered within the overall dataset.

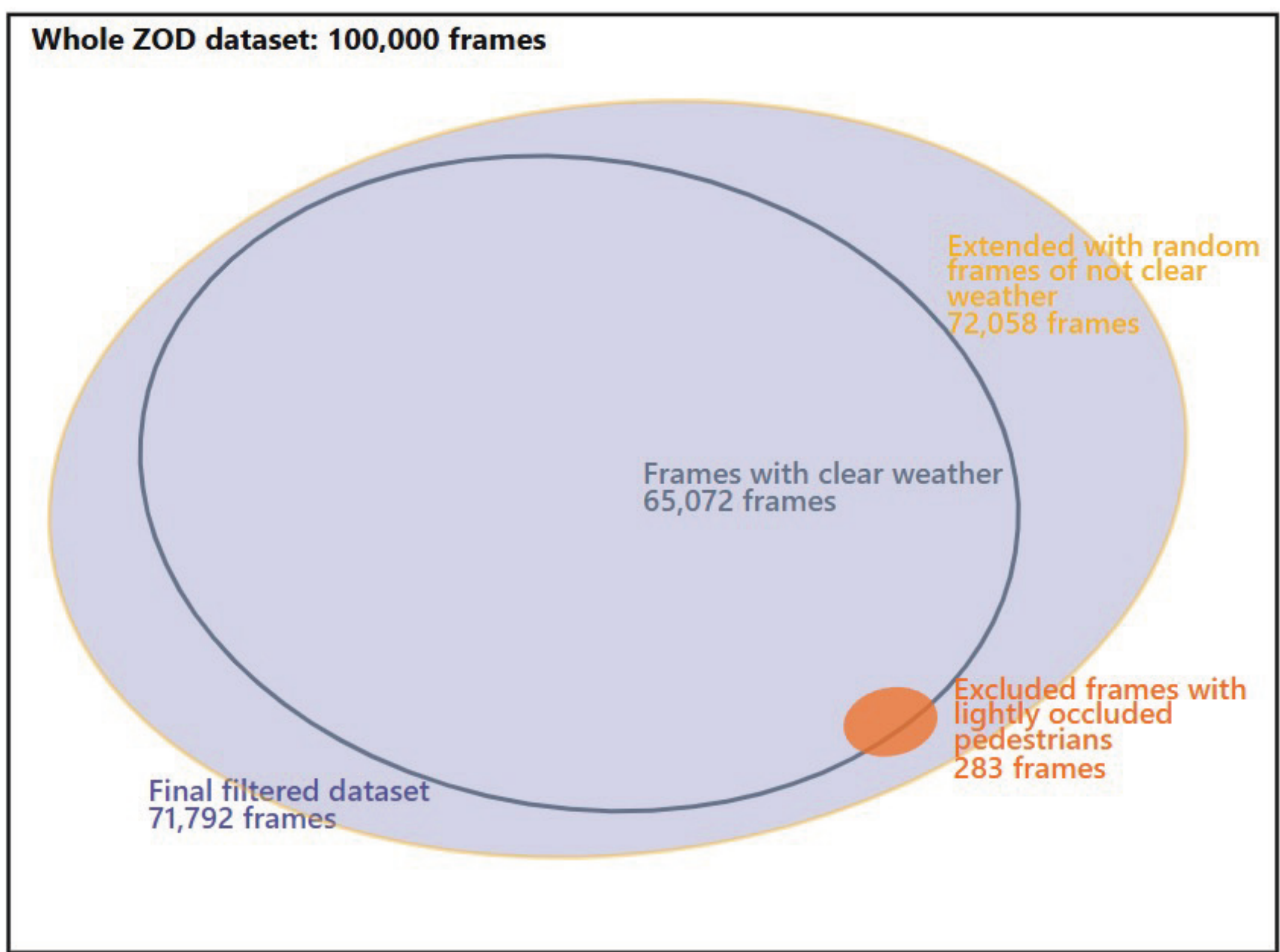


Figure 1: Post-processing for excluding specific data samples (© Bertrandt Group)

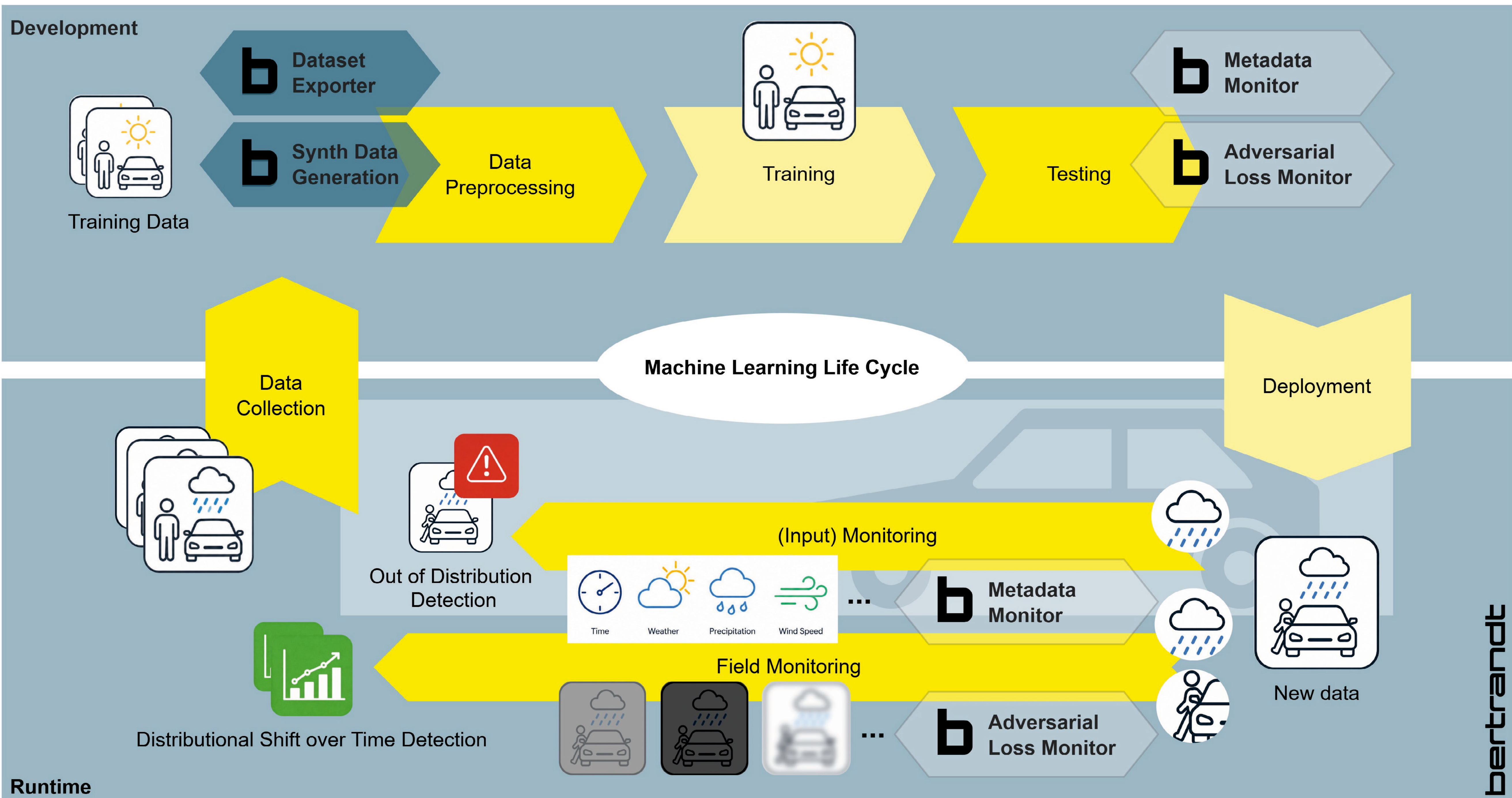


Figure 2: Machine Learning Life Cycle (© Bertrandt Group)

Synthetic Data Generation

Synthetic data generation is used to address data gaps identified during monitoring, particularly for rare weather and lighting conditions that are difficult and costly to capture in real-world operation. Realistic variants of existing camera frames are generated using a Vision Language Models with prompt-based image editing approach while preserving the original scene context.

To assess the effectiveness of the dataset, the generated samples are validated through comparative training on synthetic and real-world data. Their reliability is also assessed by analyzing whether the augmented datasets improve coverage of previously underrepresented conditions and contribute to improved model performance.

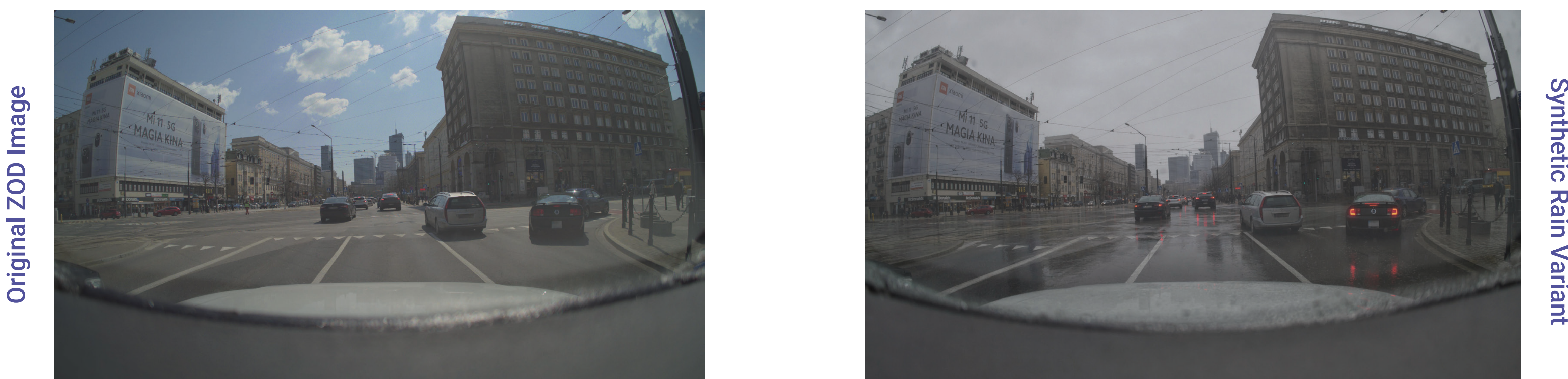


Figure 3: Original ZOD frame and GPT-image-2 synthetic rain (Original: © 2022 Zenseact AB | Synthetic modification: Bertrandt Group) *

* The images are extracted from ZOD dataset. For this dataset, Zenseact AB has taken all reasonable measures to remove all personally identifiable information, including faces and license plates. To the extent that you like to request removal of specific images from the dataset, please contact privacy@zenseact.com.

For more information regarding the poster, contact:

riken.mangukiya@bertrandt.com,
xinyi.li@bertrandt.com,
christian.freystein@bertrandt.com



Safe AI Engineering is a project of the KI Familie. It was initiated and developed by the VDA Leitinitiative autonomous and connected driving and is funded by the Federal Ministry for Economic Affairs and Energy.



[linkedin.com/company/ki-familie](https://www.linkedin.com/company/ki-familie)
www.safe-ai-engineering.de